OTS-63-21522

JPRS: 18,579

5 April 1963

CYBERNETIC ASPECTS OF STATISTICS

by Klaus-Jergen Richter

- East Germany -

Scale 4

U. S. DEPARTMENT OF COMMERCE
OFFICE OF TECHNICAL SERVICES
JOINT PUBLICATIONS RESEARCH SERVICE
Building T-30
Ohio Dr. and Independence Ave., S.W.
Washington 25, D. C.

Price: $.50

F O R E W O R D

This publication was prepared under contract for the Joint Publications Research Service as a translation or foreign-language research service to the various federal government departments.

The contents of this material in no way represent the policies, views or attitudes of the U. S. Government or of the parties to any distribution arrangement.

PROCUREMENT OF JPRS REPORTS

All JPRS reports may be ordered from the Office of Technical Services. Reports published prior to 1 February 1963 can be provided, for the most part, only in photocopy (xerox). Those published after 1 February 1963 will be provided in printed form.

Details on special subscription arrangements for JPRS social science reports will be provided upon request.

No cumulative subject index or catalog of all JPRS reports has been compiled.

All JPRS reports are listed in the Monthly Catalog of U. S. Government Publications, available on subscription at $4.50 per year ($6.00 foreign), including an annual index, from the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.

All JPRS scientific and technical reports are cataloged and subject-indexed in Technical Translations, published semimonthly by the Office of Technical Services, and also available on subscription ($12.00 per year domestic, $16.00 foreign) from the Superintendent of Documents. Semiannual indexes to Technical Translations are available at additional cost.

JPRS: 18,579

CYBERNETIC ASPECTS OF STATISTICS

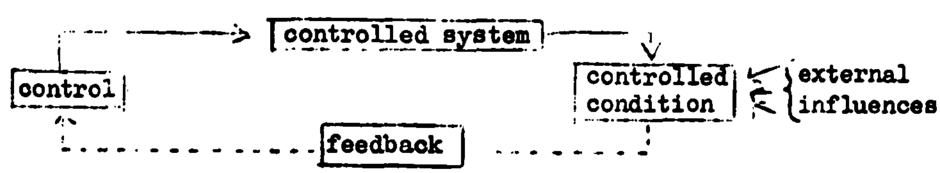
- East Germany -

[Following is a translation of an article by Klaus-Jergen Richter in Statistische Praxis (Statistical Practice), No. 1, Berlin, 19 January 1963, pages 15-19.]

The number of fields in which cybernetic systems are discovered, is constantly expanding. We need only name, for example, biology, control technics and the social sciences [Note: in regard to the latter, cf. Thiel "On Mathematical-Cybernetic Comprehension of Economic Principles" in Wirtschaftswissenschaft (Economic Science), 1962, No. 6, page 889]. In the final analysis, this concerns the recognition of systems characterized by the aspects of system, control, information and of the theory of probability (Note: cf. Klaus, in Foreward to Cybernetics by Poletajew, Berlin 1962, page XII). Such a problem formulation certainly does not exist in the search for the cybernetic aspects of statistics. Disregarding the fact that statistics plays a significant role in the theoretical structure of cybernetics, the investigation of these aspects for statistics is contained in the two questions on the role of statistics as a form of feedback and in the analogy between cybernetics and statistical categories and the type of these analogies.

Statistics as a Form of Feedback

Feedback is a decisive component of the control loop which combines the aspects of system, regulation, information and of the theory of probability:



The system aspect of the control loop is manifested in the fact that the individual components of control, controlled system, controlled condition and feedback form a cybernetic system in which information is processed. The regulatory aspect is expressed in the fact of the existence of a control and of a controlled parameter. The information aspect is manifested in the form of feedback through which information is transmitted to the control on the status of the controlled parameter. The theory of probability is that aspect of the control circuit which is expressed in the choice of action by the control for maintaining the controlled parameter at a given magnitude or value in spite of external influences.

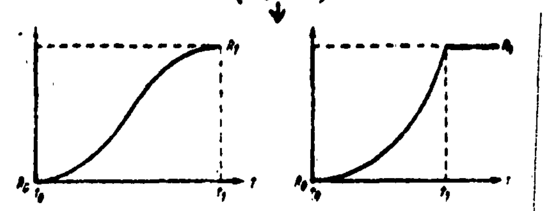
Control loops of this type exist in many forms in technics as well as in natural organisms. Technical devices can exercise the regulatory function quite as well as components of natural organisms. There are, however, also control loops in which man assumes the role of the control. Such a control loop consists for instance of a motor vehicle and its operator. This system poses regulatory tasks both in regard to the direction to be maintained by the vehicle as well as in regard to the speed of the vehicle. However, man acts as control, individually and collectively, also in larger and more complicated cybernetic systems.

There is consequently nothing which contradicts the interpretation of the national economy as a very complicated cybernetic system. Here the parameter to be controlled may be production. On the other hand, the extent of investments, accumulation or individual consumption may constitute the controlled parameter. The correct magnitude of these parameters is established by the national-economy plans in a socialist planned economy. The management organs of the national economy, the control of the system, must ascertain that the magnitude of the controlled parameter prescribed in the plan is reached and/or maintained. Their regulatory influence on the national economy is subordinate to this goal. Correct regulation, however, also requires an exact knowledge of the controlled parameter. In such an extraordinarily complicated cybernetic system like the national economy, a large number of random influences are present which are responsible for the fact that the regulatory measures transmitted by the control produce only an approximation (greater or lesser) of the status of the controlled parameter desired by the control. Furthermore, the interference in the transmission of information (the so-called noise) and the inertia (greater or lesser) with which the controlled parameter reacts to the instructions of the control, reduce the immediate and direct effectiveness of the instructions of the control. The way and manner in which the controlled parameter reacts to the control determines the behavior of the entire system. /See Note/ (Cf. Illustrative Control Technology, Berlin, 1960, page 18).

[Note]: In technical controls, it is generally relatively easy to determine the transfer functions by measurement. In the field of the social sciences, this creates a rather extensive tasks of statistics, not always easily solved, for determining the transfer functions through suitable methods. For example, let us assume that the magnitude to be controlled is the sale of certain products. We are assuming that sales do not reach the desired volume so that the control attempts to influence sales through lowering of price. In this case, the controlled system represents the price level which is varied by the control at the instant t_0. The controlled condition, the sale of the respective product, will not immediately reach the sales volume corresponding to the new price level by reason of the different causes mentioned above but assume this volume only at the instant t_1. Both the interval t_1-t_0 as well as the form of variation of the controlled condition in this interval of time are expressed by the transfer function. The two following curves are examples of transfer functions.

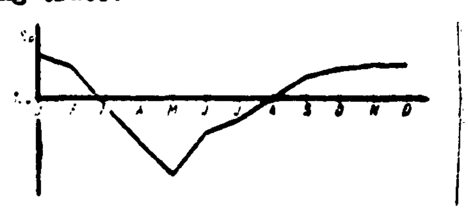
controlled condition (sales)

controlled condition (sales)



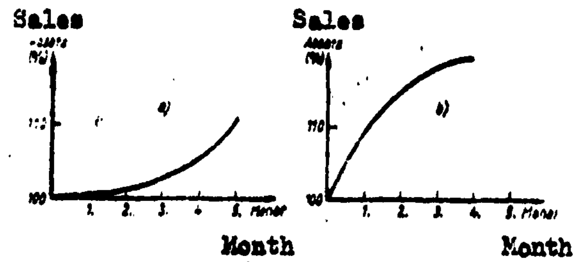
It becomes clear from the reasons mentioned that there are differences as a rule between the actual status of the controlled parameter and that desired by the control. However, effective performance by the control requires that it be continuously informed on the status of the controlled parameter. In national-economic control loops, statistics can assume the extraordinarily important function of feedback, i.e., the function of informing the control on the status of the controlled parameter.

As explanation by a simple example, let us assume that statistical data covering a certain period or time are available on the sale of certain goods. These data tend to show that the sale of the respective goods is subject to a certain annual periodicity and that moreover certain irregular variations occur. The periodic component of the time sequence shows the following trace:



We may conclude from the observed irregular variations that the effectively determined values vary on the average by a maximum of 10% above and below the norm. As shown by the sketch, sales in the month of August represent 100% of the annual average but rise in September, as a result of detailed examination of the average, to 110 and in October to 115%.

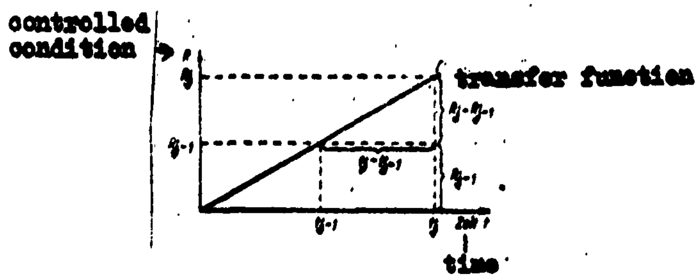
Let us assume further that prices were lowered in August to increase sales of the respective goods. If we then do not collect exact data (in the form of the transfer function mentioned above) on the manner of reaction of the controlled parameter (sales) to the instructions of the control (lowered prices), we may then be confronted by the fact that information on the status of the controlled parameter in the form of statistical data is obtained too early and that therefore no effective information can be drawn from it. From the multitude of the possible transfer functions in the case described, we shall merely select two:



In case a) no appreciable reaction of the controlled parameter occurs in the first two months after the price change. Sales increased only to a minor extent. In the third and fourth month also, the increase of sales achieved by the change of price reaches barely 10% of the starting magnitude. However, since it is known, on the basis of earlier investigations, that we must anticipate a maximum average variation of 10% from the sales volume determined, the variation of sales as consequence of lower price occurring during the first four months takes place within the limits determined by the irregular (or random) variations. A variation within these limits may be the consequence of the price change but may also be the result only of the random variations. Because an exact separation can in general be effected only with great difficulty, there is no point in attempting to procure information with the aid of (statistical) feedback during the first four months after the lowering of prices.

In case b), the transfer function shows an entirely different trace. Here the decisive changes of the controlled parameter occur immediately after the control action (after the lowering of prices). Because the variations of sales to be anticipated already in the first

months after control, lie above the limit applicable to random variations, there also exists the possibility of measuring the variation of the controlled parameter as a consequence of control action and to transmit the respective information to the control. These considerations can be generalised by means of the following sketch:



For the sake of simpli . ., a linear transfer function was assumed. However, it is now possible to introduce a magnitude

$$v = \frac{/R_j - R_{j-1}/}{t_j - t_{j-1}} \qquad (1)$$

and/or

$$v = \frac{\Delta R_j}{\Delta t_j} \qquad (1a)$$

Of course, there is always valid here

$$t_j > t_{j-1}$$
$$t_j - t_{j-1} = \frac{/R_j - R_{j-1}/}{v} \qquad (2)$$

Because t_j symbolizes a later point in time in relation to t_{j-1}, R_j and R_{j-1} indicate the magnitude of the controlled parameter R at the points in time t_j and t_{j-1}, and the magnitude v indicates the rice of the transfer function. In our sketch, v is constant. However, nothing is changed in this indication of magnitude if we admit that v itself varies. This occurs in the non-linear transfer functions which must be expected in general. From equation 1, there follows

$$t_j - t_{j-1} = \frac{/R_j - R_{j-1}/}{v} \qquad (3)$$

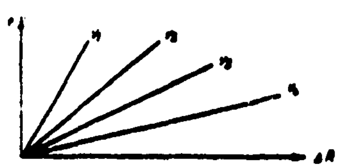
$$t_j = t_{j-1} + \frac{1}{v} \cdot \Delta R_j$$

In relation 3, t_{j-1} signifies either the instant of control action or that of the last measurement of the controlled parameter. t_j further represents the next measuring instant and v the rise of the transfer function. The relation also contains the magnitude ΔR_j, i.e., the change of the controlled parameter in the transfer from instant t_{j-1} to instant t_j.

There now exists the possibility of determining, when the minimum necessary magnitude of the change of R (i.e. ΔR_j) and the starting instant t_{j-1} is given, the instant t_j at which new information on the controlled parameter in the sense of the given formulation is of the earliest usefulness. Further simplification is possible by setting $t_{j-1} = 0$; then

$$t = \frac{1}{v} \Delta R \qquad (4)$$

expresses the interval of time in which, starting from any given instant, a new (or the first) measurement and consequent information is pertinent. This simplification is possible, however, only with such transfer functions or function components which are linear or can be regarded as linear with adequate accuracy so that constant values of rise v exist. The relation 4 can be very simply represented graphically:



For non-linear transition functions, we must turn back to relation 3. Here we know that v is then no longer constant but itself dependent on t. The investigations connected with this question are regarded by us as necessary and useful.

The above considerations were directed to the question when statistical information on the controlled parameter is of use in relation to the given transfer function. However, it is also possible to arrive at conclusions on the form of procuring statistics to be selected, with the aid of the transition function. As comparator for our decision, we utilize here the expenditure of time required for the procurement, processing and transmission of a statistical element. For the process "j," this expenditure amounts to τ_j per element. The total expenditure (of time) therefore amounts with the statistical treatment of n elements with the process "j":

$$T_j = n \cdot \tau_j \tag{5}$$

However, it is obvious that the magnitude T_j must be smaller than the interval Δt_j between the measurements of j-1 and j (statistical survey), i.e., there is valid

$$T_j < \Delta t_i \tag{6}$$

In the opposite case, the resultsof the measurement of j-1 are not yet available at the instant of the measurement of j. This means that the statistical information and its feedback produced by them lose their meaning for the activity of the control.

If we assume further that a certain amount of time is necessary for the control to arrive at decisions on the basis of the information transmitted, i.e., to react to the feedback, then relation 6 becomes

$$T_j + t_r < \Delta t_i \tag{7}$$

in which t_r represents the reaction time of the control in regard to the information obtained through feedback (Note: The magnitude t_r is assumed as independent from the process of survey. It is possible, however, that it is influenced by the type of the controlled magnitude).

Because t_j contains the number of elements n which can be predetermined because the spot-check error d (in the determination of which n enters) must be smaller than the difference ΔR_j of the controlled condition (Note: Through the assumption of the

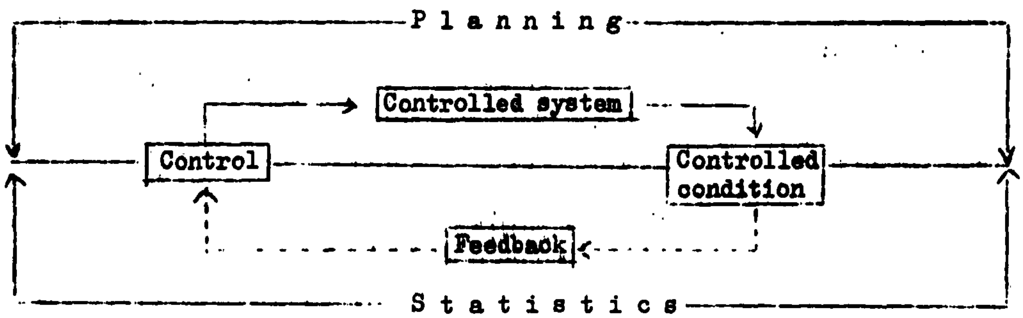
difference ΔR_j which must be determined for reasons of the control action, the extent of the statistical survey is obviously interfered with. This extent n must be determined so that the spot-check error linked to it is always smaller, when taking into account the other magnitudes influencing this error, than the smallest practically usable value of ΔR_j, indications for the magnitude τ_j can be obtained from relation 7 and, with their help, the respective method can be selected. However, we should stress that these considerations are primarily theoretical because the required standard time values τ_j for the individual statistical methods are not yet available and are not simple to determine. This is true, however, also for the transfer functions listed above. In spite of this, we believe theoretical investigations of this kind to be necessary in order to investigate the relations existing between cybernetics and statistics.

In regard to this problem complex, let us mention in conclusion that relation 7 can be further narrowed down by a condition which is related to the survey costs. It may be required not only that relation 7 is not exceeded but that simultaneously the costs connected with the survey

$$K = n \cdot k_j \longrightarrow \text{Minimum} \qquad (8)$$

assume a minimum value. Here K = total costs; n = number of surveyed elements; k_j = survey costs per element when utilizing method "j".

Equation 8 is a restricting condition to relation 7. It has been indicated what considerations arise from the possibility of interpreting statistics in the system of the socialistic national economy as a form of feedback which represents an integral component of the cybernetic system "national economy." The briefly sketched considerations may deserve to be pursued further. They illuminate the problem, frequently discussed in the past, of the interrelation between planning and statistics from the viewpoint of cybernetic concepts. Planning and statistics are consequently two components of a system



The distinction of information-processing, established by Potthoff, into ex-post-facto and ante-facto processing cannot be directly accepted on the basis of these considerations, at least in regard to planning and statistics (Note: Cf. Potthoff "Thoughts on Information Processing" in Wirtschaftswissenschaft, 1962, No. 6, page 921). Potthoff lists statistics as an example of ex-post-facto information-processing and planning as example of ante-facto processing. For ex-post-facto processing, he considers slow-working data-processing methods as adequate but feels that ante-facto processing requires rapidly operating devices (e.g. electronic computers). Actually, statistics has the character both of ante-facto as well as of ex-post-facto processing of information. In the form of feedback in the national economy, this double character of information-processing manifests itself clearly because information furnished by feedback is needed for renewed planning. In this regard, statistics is a component of a control loop, quite independently from the fact whether its information is to be furnished over longer or shorter intervals of time. Simultaneously, statistics also has the nature of ex-post-facto information processing and therefore of a chain of information. However, the question whether slow or rapidly operating data-processing installations are required, does not depend primarily on the distinction between ante-facto and ex-post-facto data-processing, especially since this distinction is not always easy to make in statistics. Statistics are perfectly able to operate with slowly working installations (e.g. punch-card systems) when the changes in the system itself take place slowly and the minimum magnitude of the changes of the controlled condition ΔR_j becomes adjusted only very slowly. At the same time, it maintained that the required calculating operations can be effected by punch-card systems. Although this specifically concerns work which has the nature of ante-facto information-processing, the utilization of expensive and quickly operating electronic devices for data-processing would often be uneconomical in these cases.

Are There Analogies Between Statistical and Cybernetic Categories?

As in the discussion of the question on the role of statistics as a form of feedback, the following remarks will be able to only touch on the question. For one thing, there do not as yet exist sufficient investigations of this problem complex and it will also require collective efforts in order to arrive at further results. The analogy between statistical and cybernetic categories refers primarily, as far as we are now able to determine, to relations between the theory of information (as a component of cybernetics) and statistics. Adam (Note: Cf. Adam, "The Maximum Determination Scale in the Case of Relations Between Qualitative Characteristics" in The Utilization of Matrix Calculation to Economical and Statistical Problems, Wuerzburg 1959, page 210) has published several communications on this and has treated in particular the relation between entropy and scatter (standard deviation) and questions of dependency measurement by utilizing considerations from the theory of information. The opinion expressed by Adam that statistics tends to a general science of information (Note: Cf. Adam, "Entropy and Scatter" in Metrica 1958-1959, page 99) merits serious examination. It would seem necessary to undertake investigations in the field of theoretical statistics in the future also from the viewpoint of the cybernetic aspects of statistics. This will very likely produce both theoretical as well as practical benefit.

The following will attempt to clarify the question here investigated through the example of a comparison between the cybernetic category of being organized and the statistical concept of correlation (the remarks on theory of information refer to Poletajew, Ibid., page 89).

The entropy of the number of states of the element X_j of a system is known to be

$$H_i(X_i) = - \sum_k P_{ik} \cdot \log P_{ik} \qquad (9)$$

in which n possible states of the element X_j $(k = 1, 2, \ldots n)$ are assumed. P_{jk} is the probability that the element X_j will be specifically in the state k. It is further assumed that the system consists of the two elements X_1 and X_2. If these elements are independent of each other, the total entropy of the system is

$$H_o(X_1X_2) = H(X_1) + H(X_2) \qquad (10)$$

i.e., it is equal to the sum of the individual entropies. In this case, we speak of a maximum non-organized system. If the maximum non-organization is eliminated and connections and/or relations have been created between the elements X_1 and X_2 in the system, then the total entropy of the more or less organized system $H_1(X_1/X_2)$ is smaller than the maximum entropy H_0 which exists only when non-organized. There is then valid

$$H_1(X_1X_2) = H(X_1) + H(X_2/X_1) \leq H(X_1) + H(X_2) \quad (11)$$

$$= H(X_2) + H(X_1/X_2) \leq H(X_1) + H(X_1) \quad (12)$$

Here $H(X_1/X_2)$ and $H(X_2/X_1)$ represent conditional entropies, i.e., in the first case, in regard to the selection of X_1 if X_2 is known and, in the second, in regard to the selection of X_2 if X_1 is known. Since we proceeded from the assumption that relations exist between two elements of the system and that they are therefore not independent of each other, the uncertainty in regard to the selection of one element must decrease if the other element related to it is already known (Note: A very simple example for this are piece-work wages in which the wages depend on output. The uncertainty in regard to the selection of an employee with a certain wage is appreciably reduced by knowledge of his output. (This also points to relations between the concept of entropy and the problem of selection of the random-sample method)). It follows from this that

$$H(X_1/X_2) < H(X_1) \quad (13a)$$

$$H(X_2/X_1) < H(X_2) \quad (13b)$$

This proves the accuracy of equations 11 and 12. As a quantitative measure of the stage of organization reached between X_1 and X_2, the difference between maximum entropy H_0 and entropy H_1 (reduced by organization) can be used in the form of

$$\Delta H = H_0 - H_1 \quad (14)$$

and/or

$$\Delta H = H(X_1) - H(X_1/X_2) \quad (15a)$$

and

$$\Delta H = H(X\ ) - H(X\ /X\ ) \quad (15b)$$

This magnitude is equal to zero when no degree of organization exists because then the condition is equal to the non-conditional entropy and the knowledge of the one element does not reduce the entropy existing in regard to the selection of the other. Both elements are then not interrelated. In greater or lesser organization of the system, ΔH is not equal to zero. It would seem pertinent to express the difference ΔH relatively. For equation 15a, we then obtain

$$\Delta H_r = \frac{H(X_1) - H(X_1/X_2)}{H(X_1)}$$

and/or

$$\Delta H_r = 1 - \frac{H(X_1/X_2)}{H(X_1)} \qquad (16)$$

Since conditional and non-conditional entropy concord in the complete absence of organization, we have

$$\Delta H_r = 0$$

On the other hand, the conditional entropy $H(X_1/X_2)$ cannot become less than zero. If a complete, i.e., functional, interrelation exists between the two elements so that knowledge of X_2 completely eliminates the uncertainty existing in regard to the selection of X_1, we then have

$$\Delta H_r = 1$$

Accordingly, we can then delimit the relative entropy difference in the form

$$0 \leqq \Delta H_r \leqq 1 \qquad (17)$$

and now are confronted by ΔH_r by a magnitude which corresponds to the scale of certainty (or square of the correlation coefficient $D = r^2$) familiar in statistics. For B, there is also valid

$$0 \leqq B \leqq 1 \qquad (18)$$

in which the case "zero" corresponds to the state of complete independence and the case "one" corresponds to the functional dependence between the two statistical measurement series.

The cases of dependence between two statistical series of characteristics to be treated by the theory of correlation now

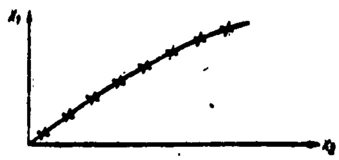
reveal themselves to be a special case of the more general cybernetic concept of organization. This fact also becomes clear when we regard each of the two elements here mentioned as a numerical interval on a straight line with a frequency distribution given above the latter as is customary in correlation calculation. The following sketches are therefore also completely identical with those familiar from correlation calculation (Note: Cf. Poletajew, Ibid., page 90).

Frequency Distribution of the States of Two Elements X_1 and X_2:

a) Under complete absence of organization ($\Delta H_r = 0$; $B = 0$)

b) When organized different from zero ($0 < H_r < 1$; $0 < B < 1$)

c) Under maximum organization ($\Delta H_r = 1$; $B = 1$)



To discuss in detail the various consequences of the analogy just demonstrated between scale of organization and scale of certainty, would exceed the scope of this article. We may be sure, however, that the utilization of cybernetic concepts is able to contribute to the clarification and understanding of statistical categories. Undoubtedly, others besides the analogies here given will be found. However, that utilization of cybernetics in statistics will produce not only theoretical but also practical benefit for statistical work, was to be demonstrated by means of the remarks made in the section on statistics as a form of feedback.

(The observations contained in this article were delivered as a lecture at the Karl Marx Institute of Economic Sciences in Sofia on 15 November 1962).

6198
CSO: 1872-D

\- END -